M2R Maths Applis-INSA GMM
2010–2011
Modélisation Statistique

TP du 5 Janvier 2011 de 9h30 à 12h30

1 Choix de Modèle

On considère le modèle de régression simple :

$$Y_i := \sum_{j=1}^{10} \theta_j^* x_{ij} + \varepsilon_i,\ i = 1 \dots N.$$

Où $\varepsilon_i,\ i = 1 \dots N$ sont des gaussiennes centrées indépendantes et de même variance σ_*^2. On générera ce modèle en utilisant comme regresseurs (x_{ij}) des réalisations indépendantes de variables uniformes sur $[-1, 1]$ qui seront fixées une bonne fois pour toute (générer ces variables avec $N = 750$ puis sauver). On se placera dans le cas où très peu des paramètres θ_j^* sont non nuls (prendre juste 2, 3 ou 4 paramètres non nuls). Mettre en place les techniques de choix de modèles étudiées en cours pour trouver le *meilleur* modèle. On rappelle les différents critères :

- C_p-Mallows

$$C_p(m) = \frac{\widehat{\sigma^2_{(m)}}}{\widehat{\sigma^2}} + 2\frac{|m|}{N},$$

- AIC corrigé

$$AIC(m) = \log\left(\widehat{\sigma^2_{(m)}}\right) + \frac{N + |m| + 1}{N - |m| - 3},$$

- BIC

$$AIC(m) = \log\left(\widehat{\sigma^2_{(m)}}\right) + \frac{\log N}{N}|m|.$$

Afin d'explorer facilement les 1024 modèles on construira une liste aléatoire de $1024[\log 1024]$ modèles obtenue par la stratégie du *collectionneur!!*.

2 Conditionnement et Processus gaussiens

1. Construire à l'aide de variables gaussiennes i.i.d. $\mathcal{N}(5, 5^2)$ une matrice de covariance Γ de taille 10×10 et un vecteur m de taille 10. Stocker Γ et m. Simuler N réalisations de la loi $\mathcal{N}_{10}(m, \Gamma)$. Calculer la prédiction de la première composante quand on connaît les 9 dernières. Comparer à la loi conditionnelle théorique.
2. Simuler une trajectoire d'un processus $AR(1)$ gaussiens calculer les prédicteurs à horizon, $1 \dots, d > 1$ et comparer à la valeur du processus.
3. Télécharger la boîte à outils *DACE*. Utiliser la boîte à outil pour approximer une fonction donnée à 1 puis 2 variables.